

MINT

Morteza Sabri¹

¹National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

2021-05-02 (updated: 2021-05-04)

Abstract

MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms

Contents

1	Datasets	2
1.1	Download	2
1.2	Deleting Gene Symbols with no character or NAs	3
1.3	All exprs into one list	4
1.4	Annotation	4
1.5	Deleting probe ids with multiple gene id	4
1.6	Deleting Duplicated Gene Symbols	5
1.7	Concatenate	5
1.8	Grouping, Study and Tissue	5
1.9	raw boxplot.	5
2	MINT	6
2.1	Selected Transcripts	6
2.2	Sample Plot	7
2.3	Correlation Circle Plot	7
2.4	Clustered Image Map.	9
2.5	Relevance Network.	9
2.6	Loading Plot	10
2.7	ROC	11
	Session info	11

1 Datasets

- GSE74602
- GSE110223
- GSE110224

1.1 Download

```
## $GSE74602
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 22184 features, 60 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1923670 GSM1923671 ... GSM1923729 (60 total)
##   varLabels: title geo_accession ... tissue:chl (34 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: ILMN_1343291 ILMN_1343292 ... ILMN_2038778 (22184 total)
##   fvarLabels: ID Gene title ... Platform_SEQUENCE (22 total)
##   fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## Annotation: GPL6104
##
## $GSE110223
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 22283 features, 26 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM2982906 GSM2982907 ... GSM2982931 (26 total)
##   varLabels: title geo_accession ... vb pet variable values:chl (64 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: 1007_s_at 1053_at ... AFFX-TrpnX-M_at (22283 total)
##   fvarLabels: ID Gene title ... GO:Component ID (21 total)
##   fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
##   pubMedIds: 30809322
## Annotation: GPL96
##
## $GSE110224
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 54675 features, 34 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM2982932 GSM2982933 ... GSM2982965 (34 total)
##   varLabels: title geo_accession ... vb pet variable values:chl (63 total)
##   varMetadata: labelDescription
## featureData
```

```
## featureNames: 1007_s_at 1053_at ... AFFX-TrpnX-M_at (54675 total)
## fvarLabels: ID Gene title ... G0:Component ID (21 total)
## fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## pubMedIds: 30809322
## Annotation: GPL570
```

1.2 Deleting Gene Symbols with no character or NAs

```
## $GSE74602
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 20498 features, 60 samples
## element names: exprs
## protocolData: none
## phenoData
## sampleNames: GSM1923670 GSM1923671 ... GSM1923729 (60 total)
## varLabels: title geo_accession ... tissue:chl (34 total)
## varMetadata: labelDescription
## featureData
## featureNames: ILMN_1343291 ILMN_1343292 ... ILMN_2038778 (20498 total)
## fvarLabels: ID Gene title ... Platform_SEQUENCE (22 total)
## fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## Annotation: GPL6104
##
## $GSE110223
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 21154 features, 26 samples
## element names: exprs
## protocolData: none
## phenoData
## sampleNames: GSM2982906 GSM2982907 ... GSM2982931 (26 total)
## varLabels: title geo_accession ... vb pet variable values:chl (64 total)
## varMetadata: labelDescription
## featureData
## featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at (21154 total)
## fvarLabels: ID Gene title ... G0:Component ID (21 total)
## fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## pubMedIds: 30809322
## Annotation: GPL96
##
## $GSE110224
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45113 features, 34 samples
## element names: exprs
## protocolData: none
## phenoData
## sampleNames: GSM2982932 GSM2982933 ... GSM2982965 (34 total)
## varLabels: title geo_accession ... vb pet variable values:chl (63 total)
```

```
## varMetadata: labelDescription
## featureData
## featureNames: 1007_s_at 1053_at ... AFFX-HUMISGF3A/M97935_MB_at (45113 total)
## fvarLabels: ID Gene title ... G0:Component ID (21 total)
## fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## pubMedIds: 30809322
## Annotation: GPL570
```

1.3 All exprs into one list

```
## List of 3
## $ GSE74602 : num [1:20498, 1:60] 0.034 0.7627 0.5062 -0.0427 -0.0923 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:20498] "ILMN_1343291" "ILMN_1343292" "ILMN_1343293" "ILMN_1343294" ...
## .. ..$ : chr [1:60] "GSM1923670" "GSM1923671" "GSM1923672" "GSM1923673" ...
## $ GSE110223: num [1:21154, 1:26] 10.98 7.12 7.03 9.54 5.18 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:21154] "1007_s_at" "1053_at" "117_at" "121_at" ...
## .. ..$ : chr [1:26] "GSM2982906" "GSM2982907" "GSM2982908" "GSM2982909" ...
## $ GSE110224: num [1:45113, 1:34] 11.71 8.61 5.89 8.21 3.69 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:45113] "1007_s_at" "1053_at" "117_at" "121_at" ...
## .. ..$ : chr [1:34] "GSM2982932" "GSM2982933" "GSM2982934" "GSM2982935" ...
```

1.4 Annotation

```
## List of 3
## $ GSE74602 : num [1:20498, 1:60] 0.034 0.7627 0.5062 -0.0427 -0.0923 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:20498] "EEF1A1" "TUBB2A" "TXN" "ACTB" ...
## .. ..$ : chr [1:60] "GSM1923670" "GSM1923671" "GSM1923672" "GSM1923673" ...
## $ GSE110223: num [1:21154, 1:26] 10.98 7.12 7.03 9.54 5.18 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:21154] "MIR4640///DDR1" "RFC2" "HSPA6" "PAX8" ...
## .. ..$ : chr [1:26] "GSM2982906" "GSM2982907" "GSM2982908" "GSM2982909" ...
## $ GSE110224: num [1:45113, 1:34] 11.71 8.61 5.89 8.21 3.69 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:45113] "MIR4640///DDR1" "RFC2" "HSPA6" "PAX8" ...
## .. ..$ : chr [1:34] "GSM2982932" "GSM2982933" "GSM2982934" "GSM2982935" ...
```

1.5 Deleting probe ids with multiple gene id

```
## List of 3
## $ GSE74602 : num [1:20498, 1:60] 0.034 0.7627 0.5062 -0.0427 -0.0923 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:20498] "EEF1A1" "TUBB2A" "TXN" "ACTB" ...
## .. ..$ : chr [1:60] "GSM1923670" "GSM1923671" "GSM1923672" "GSM1923673" ...
```

```
## $ GSE110223: num [1:19931, 1:26] 7.12 7.03 9.54 5.18 7.29 ...
## .. attr(*, "dimnames")=List of 2
## ..$ : chr [1:19931] "RFC2" "HSPA6" "PAX8" "GUCA1A" ...
## ..$ : chr [1:26] "GSM2982906" "GSM2982907" "GSM2982908" "GSM2982909" ...
## $ GSE110224: num [1:42899, 1:34] 8.61 5.89 8.21 3.69 7.45 ...
## .. attr(*, "dimnames")=List of 2
## ..$ : chr [1:42899] "RFC2" "HSPA6" "PAX8" "GUCA1A" ...
## ..$ : chr [1:34] "GSM2982932" "GSM2982933" "GSM2982934" "GSM2982935" ...
```

1.6 Deleting Duplicated Gene Symbols

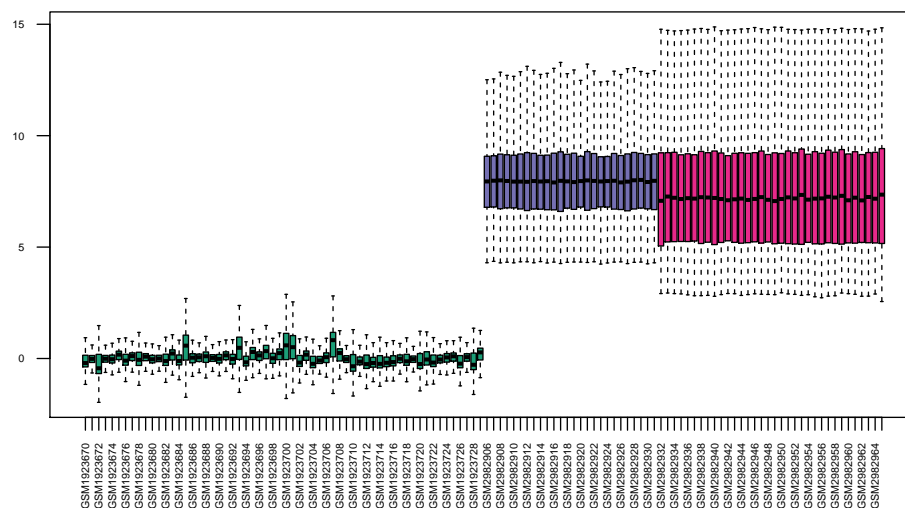
```
## List of 3
## $ GSE74602 : num [1:17181, 1:60] -0.4418 -0.3438 -0.0694 -0.3516 -0.5251 ...
## .. attr(*, "dimnames")=List of 2
## ..$ : chr [1:17181] "A1CF" "A2ML1" "A4GALT" "A4GNT" ...
## ..$ : chr [1:60] "GSM1923670" "GSM1923671" "GSM1923672" "GSM1923673" ...
## $ GSE110223: num [1:12500, 1:26] 10.28 7.77 6.38 9.44 8.34 ...
## .. attr(*, "dimnames")=List of 2
## ..$ : chr [1:12500] "A1CF" "A4GALT" "A4GNT" "AAAS" ...
## ..$ : chr [1:26] "GSM2982906" "GSM2982907" "GSM2982908" "GSM2982909" ...
## $ GSE110224: num [1:20844, 1:34] 5.14 10.24 4.69 3.92 5.7 ...
## .. attr(*, "dimnames")=List of 2
## ..$ : chr [1:20844] "A1BG-AS1" "A1CF" "A2M-AS1" "A2ML1" ...
## ..$ : chr [1:34] "GSM2982932" "GSM2982933" "GSM2982934" "GSM2982935" ...
```

1.7 Concatenate

1.8 Grouping, Study and Tissue

1.9 raw boxplot

```
## pdf
## 2
```



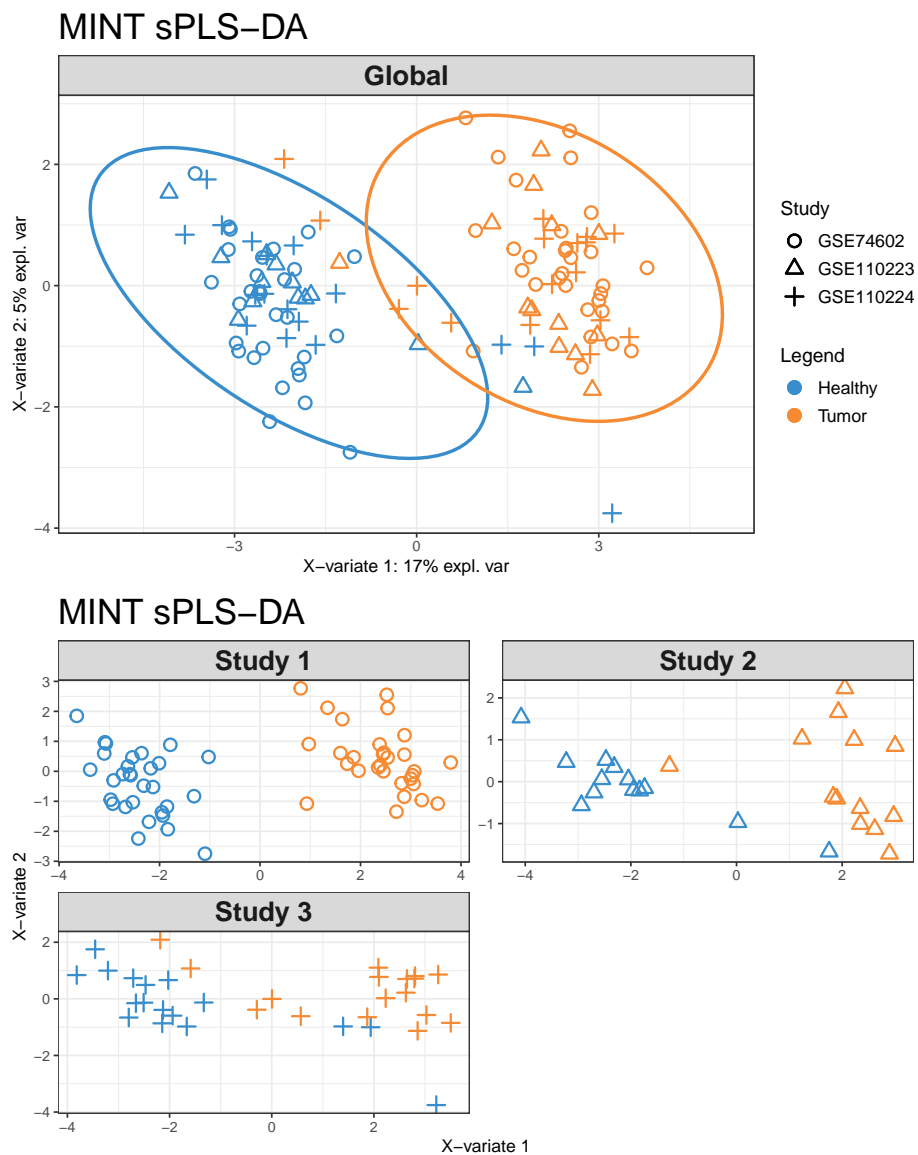
2 MINT

```
## Healthy  Tumor
##      60      60
##      study
## Y      GSE74602 GSE110223 GSE110224
## Healthy      30      13      17
## Tumor        30      13      17
```

2.1 Selected Transcripts

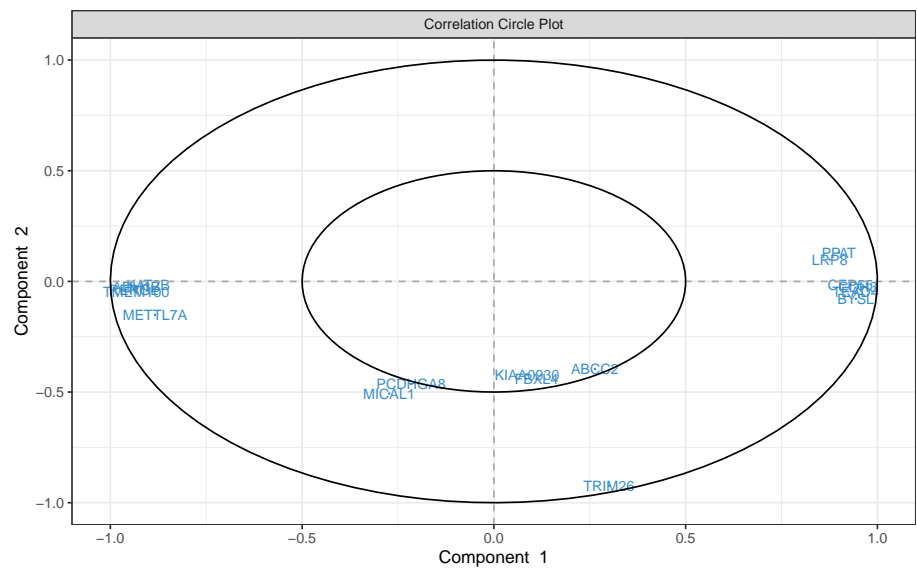
```
## $name
## [1] "CDH3"      "RERGL"      "TEAD4"      "TMEM100"    "BYSL"      "ADH1B"      "CEP55"      "METTL7A"
## [9] "KAT2B"      "PPAT"      "LRP8"
##
## $value
##      value.var
## CDH3      0.53394033
## RERGL     -0.47491565
## TEAD4      0.38928441
## TMEM100   -0.38336054
## BYSL       0.32834300
## ADH1B     -0.23884068
## CEP55      0.10682962
## METTL7A   -0.09615204
## KAT2B     -0.06942520
## PPAT       0.02172650
## LRP8       0.00717752
##
## $comp
## [1] 1
```

2.2 Sample Plot

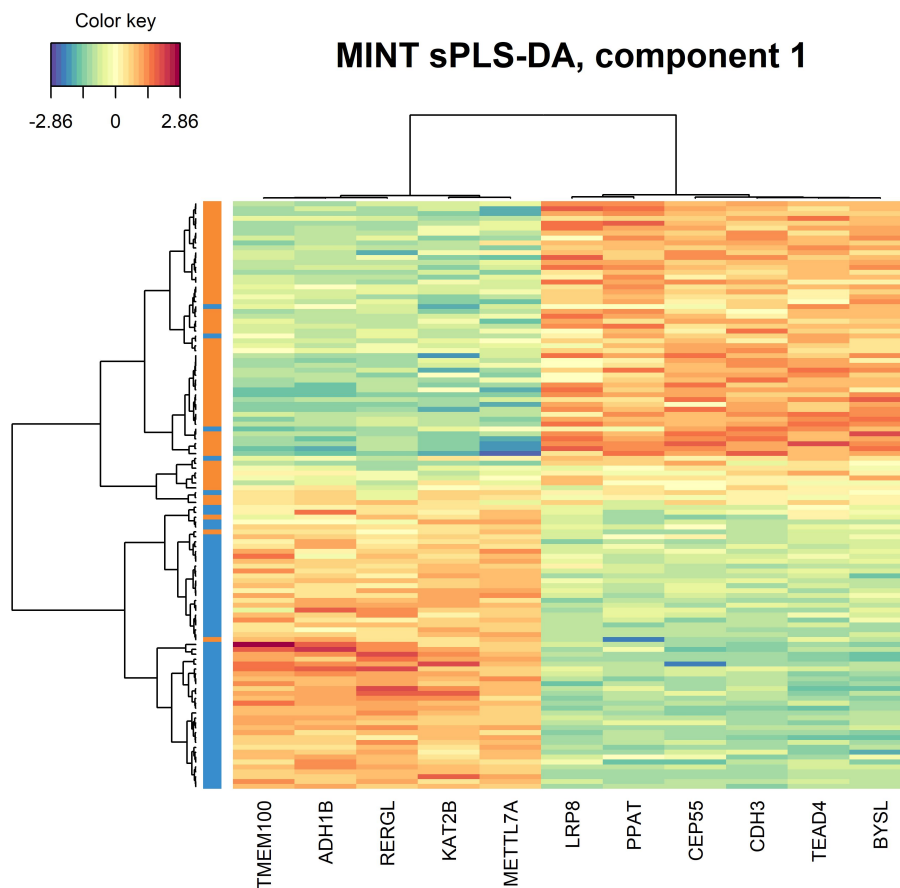


2.3 Correlation Circle Plot

highlights the contribution of each selected transcripts to each component (close to the large circle). See also more examples [here](#).



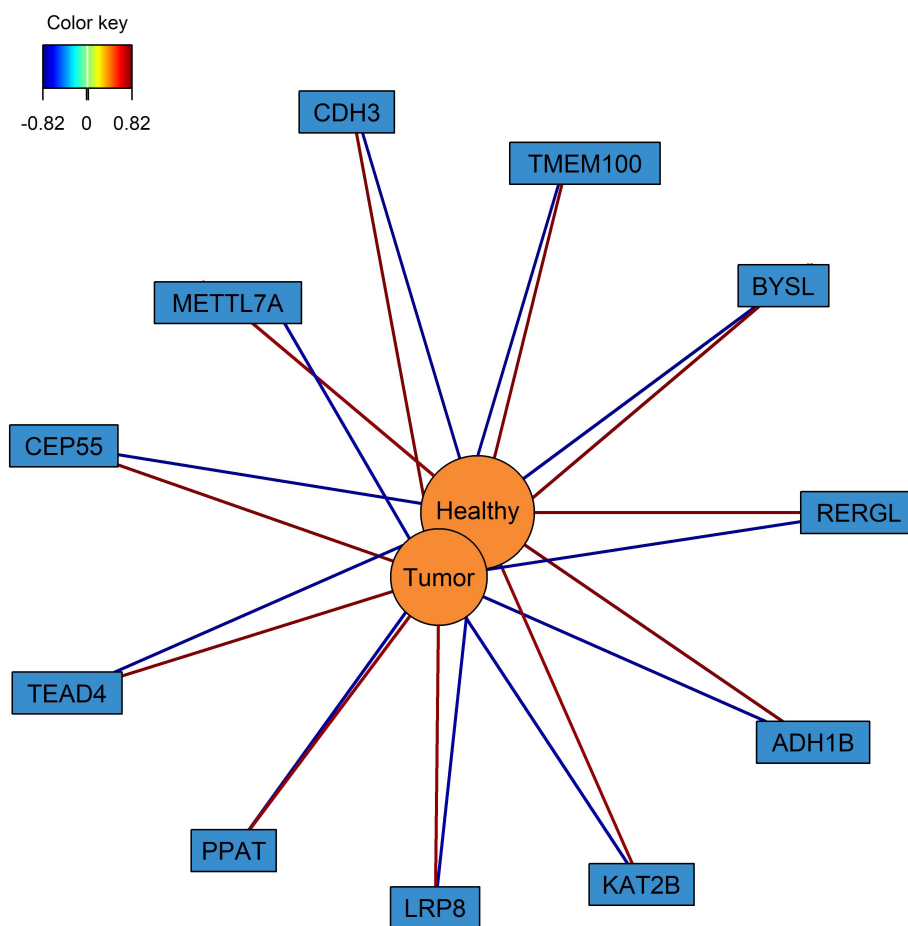
2.4 Clustered Image Map



More info about **cim** on [this link](#).

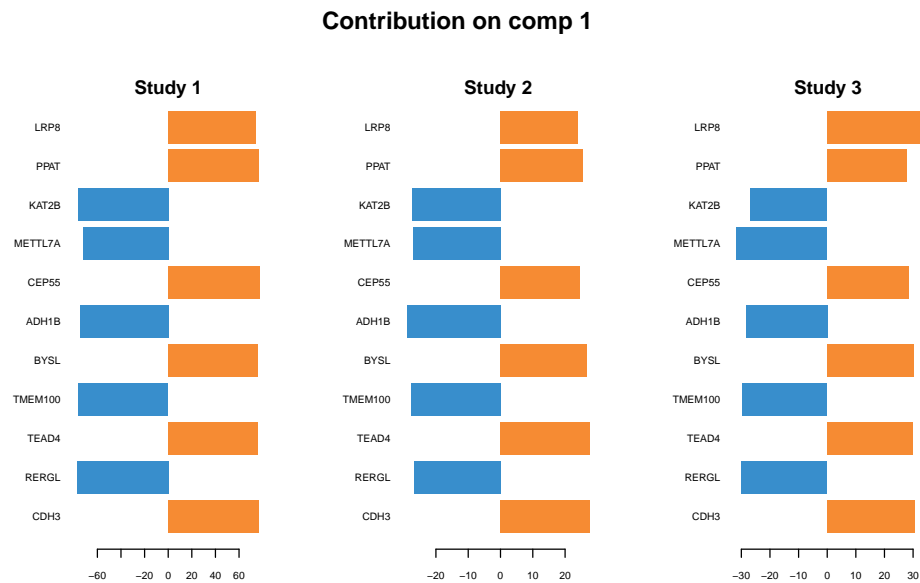
2.5 Relevance Network

Relationship between the selected variables and the outcome of each category.



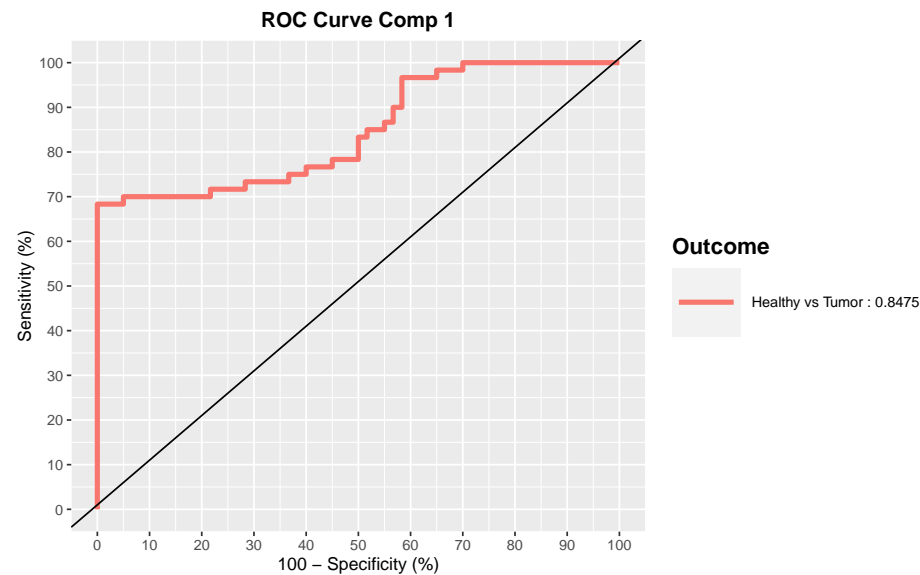
2.6 Loading Plot

Displays the coefficient weight of each selected variable in each study and shows the agreement of the gene signature across studies. Colors indicate the class in which the mean expression value of each selected gene is maximal.



2.7 ROC

Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) curves for supervised classification.



```
## $Comp1
##                               AUC   p-value
## Healthy vs Tumor 0.8475 5.166e-11
```

Session info

R version 4.0.5 (2021-03-31)

Platform: x86_64-w64-mingw32/x64 (64-bit)

locale: LC_COLLATE=English_United_States.1252, LC_CTYPE=English_United_States.1252, LC_MONETARY=English_United_States.1252, LC_NUMERIC=C and LC_TIME=English_United_States.1252

attached base packages: *parallel, stats, graphics, grDevices, utils, datasets, methods and base*

other attached packages: *mixOmics(v.6.14.1), ggplot2(v.3.3.3), lattice(v.0.20-41), MASS(v.7.3-53.1), GEOquery(v.2.58.0), Biobase(v.2.50.0), BiocGenerics(v.0.36.1) and BiocStyle(v.2.18.1)*

loaded via a namespace (and not attached): *tidyselect(v.1.1.0), xfun(v.0.22), pander(v.0.6.3), reshape2(v.1.4.4), purrr(v.0.3.4), colorspace(v.2.0-0), vctrs(v.0.3.7), generics(v.0.1.0), htmltools(v.0.5.1.1), yaml(v.2.2.1), utf8(v.1.2.1), rlang(v.0.4.11), pillar(v.1.6.0), glue(v.1.4.2), withr(v.2.4.2), DBI(v.1.1.1), BiocParallel(v.1.24.1), RColorBrewer(v.1.1-2), plyr(v.1.8.6), matrixStats(v.0.58.0), lifecycle(v.1.0.0), stringr(v.1.4.0), munsell(v.0.5.0), gtable(v.0.3.0), evaluate(v.0.14), labeling(v.0.4.2), knitr(v.1.33), curl(v.4.3), fansi(v.0.4.2), rARPACK(v.0.11-0), Rcpp(v.1.0.6), readr(v.1.4.0), corpcor(v.1.6.9), scales(v.1.1.1), BiocManager(v.1.30.12), limma(v.3.46.0), farver(v.2.1.0), RSpectra(v.0.16-0), gridExtra(v.2.3), ellipse(v.0.4.2), hms(v.1.0.0), digest(v.0.6.27), stringi(v.1.5.3), bookdown(v.0.22), dplyr(v.1.0.5), ggrepel(v.0.9.1), grid(v.4.0.5), cli(v.2.5.0), tools(v.4.0.5), magrittr(v.2.0.1), tibble(v.3.1.1), crayon(v.1.4.1), tidyr(v.1.1.3), pkgconfig(v.2.0.3), Matrix(v.1.2-18), ellipsis(v.0.3.1), xml2(v.1.3.2), assertthat(v.0.2.1), rmarkdown(v.2.7), rstudioapi(v.0.13), R6(v.2.5.0), igraph(v.1.2.6) and compiler(v.4.0.5)*